3D face reconstruction in a binocular passive stereoscopic system using face properties

Amel Aissaoui, Jean Martinet, Chaabane Djeraba

3D FACE RECONSTRUCTION IN A BINOCULAR PASSIVE STEREOSCOPIC SYSTEM USING FACE PROPERTIES

Amel AISSAOUI, Jean MARTINET and Chaabane DJERABA

LIFL UMR Lille1-CNRS n 8022, IRCICA, 50 avenue Halley, 59658 Villeneuve d'Ascq, France

ABSTRACT

In this paper, we introduce a novel approach for face stereo reconstruction in passive stereo vision system. Our approach is based on the generation of a facial disparity map, requiring neither expensive devices nor generic face models. It consists of incorporating face properties in the disparity estimation to enhance the 3D face reconstruction in real time application. An algorithm based on the Active Shape Model (ASM) is proposed to acquire 3D sparse estimation of the face with a high confidence. Using sparse estimation as guidance and considering the face symmetry and smoothness, the dense disparity is completed. Experimental results are presented to demonstrate the reconstruction accuracy of the proposed method.

***Index Terms*—** 3D reconstruction, binocular stereovision, Face properties.

1. INTRODUCTION

Depth estimation for faces is an important problem that has been conjointly studied with face animation [1], facial analysis and face recognition [2].

In the past few decades, many approaches have been proposed, including 3D from stereo matching [3], 3D morphable model based methods [4], structure from motion [5] and shape from shading techniques [6]. However, how to efficiently acquire facial depth information from stereo images is still a challenging problem, especially in binocular passive systems, where only one image pair is used and no structural lighting is available.

In the literature, a wide spectrum of works dealing with the problem of 3D reconstruction using a binocular passive stereo systems is proposed [7]. The major problem of these approaches lies in the definition of a stereo matching scheme for a given image pair. Indeed, this problem is more crucial with poorly-textured face images. This leads to ambiguities and additional complexities for matching algorithms. Therefore, a few approaches for dense 3D face reconstruction in binocular passive stereo system have been proposed comparing to those of active and multi-view stereo vision. Most of the existing methods are based on a fitting step of the estimated depth to a generic 3D model [8, 9, 10]. Le et al. [8] have built a coarse shape estimation based on 3D key points, and then used a linear morphable model to efficiently match the detailed shape and texture. Authors in [9] fit a sparse reconstruction of manually selected points to a generic model using a Thin-Plate Spline (TPS) method. In [10], a reference 3D face is used as an intermedium for correspondence calculation. The virtual face images with known correspondences are first synthesized from the reference face. Then the known correspondences are extended to the incoming stereo face images, using face alignment and warping. The complete 3D face can thus reliably be reconstructed from stereo images. The problems of these methods are the high cost of processing time related to the fitting step and the manual initialization requirement e.g. in [10], for the face alignment step. Another disadvantage of these methods is the fact that the resulting faces are more similar to the generic model than to their specific model. In [11], Lengagne et al. proposed to apply an iterative algorithm to refine the face model resulting from the fitting process of the sparse disparity to the 3D generic model using differentials constraints. This method has an additional computation cost since it includes an iterative deformation step plus the calculation of the principle curvatures on each vertex. Also, it is very sensitive to noise because it uses the second derivative for calculating curves. Some attempts have been proposed to use the Shape From Shading (SFS) method to enhance the stereo matching process. Cryer et al. [12] proposed to merge in the frequency domain the dense depth maps obtained separately from stereo and from SFS. However, in this method both processes are very sensitive to the lighting conditions. Besides, for SFS, it is assumed that the surface reflectance map or at least its form is known.

In this paper, we propose an improved method for determining the disparity information of a human face from stereo matching in a binocular vision system without using neither 3D generic model nor iterative process. Our method consists of incorporating smoothness, symmetry, and topological face properties the estimation process to improve the result of the matching process.

The remainder of this paper is organized as follows. First, we describe the depth estimation process in Section 2. In Section 3, we introduce the proposed method. In Section 4, we show preliminary results of this ongoing work, whose goal is to build 3D face models using entirely passive techniques. Finally, Section 5 concludes the paper.

2. CORRELATION-BASED METHOD FOR DEPTH ESTIMATION

Depth estimation process in binocular passive stereo system consists of reconstruction of 3D information of a scene captured from two different points of view. A step called *disparity estimation* is crucial for the reconstruction of the depth information. In order to estimate a disparity map for the scene, it is necessary to find pixels in both images that correspond to the projection of the same real word point. This process is called *stereo matching* process. While many stereo matching algorithms have been proposed [7], correlation-based algorithms still have an edge due to speed and less memory requirements [13].

A correlation-based stereo matching algorithms typically produces dense depth maps by calculating the matching costs for each pixel at each disparity level in a certain range. Afterwards, the matching costs for all disparity levels can be aggregated within a certain neighborhood window. Finally, the algorithm searches for the lowest cost match for each pixel. The most common is a simple winner takes-all (WTA) minimum or maximum search over all possible disparity levels. Different similarity measures are used in correlation- based methods. The most common ones are: Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD), Normalized Cross Correlation (NCC) and Sum of Hamming Distances (SHD).

These methods suffer from 3 essential problems :

1. Implicit hypothesis: all points of window move with same motion, that is they are in a fronto-parallel plane.

2. Aperture problem: the context can be too small in certain regions, lack of information.

3. Adherence problem: intensity discontinuities influence strongly the estimated disparity and if it corresponds with a depth discontinuity, we have a tendency to dilate the front object.

Assuming the rigidity of the face and its smoothness, the first and the third problems will not influence the estimation result. However, the aperture problem is a major problem for face disparity estimation since the face has many homogenous areas which can generate holes, spikes and many uncertain disparities. We propose to overcome the aperture problem by incorporating the smoothness, the symmetry and the topological properties of the face while maintaining its real-time suitability.

After the stereo matching process, depth information of a point $p(x, y, z)$ with an estimated disparity d is done as:

$$d = \frac{fb}{z}. \tag{1}$$

where f, b are the camera focal and baseline respectively.

3. PROPOSED METHOD FOR DISPARITY ESTIMATION

In order to estimate the disparity map, photo-consistency measures used in correlation-based method are not always sufficient to recover precise geometry, particularly in low-textured scene regions (the aperture problem). It can therefore be helpful to incorporate face proprieties that bias the reconstruction to have desired characteristics.

The disparity estimation is done in two steps: first, we calculate the sparse disparity of the face points with high confidence. Then, using this result, dense disparity map for the whole face is estimated using the common correlation method considering the topological information, the smoothness and the symmetry properties of the face.

3.1. Sparse disparity calculation

In order to establish the sparse disparity, we apply an ASM (Active Shape Model) [14] fitting algorithm on both images instead of correlation method. We use the ASM fitting, because in addition of the color information used in correlation methods, it includes the shape information obtained by an off-line learning process on a face database, which guarantees a good face features localisation in the stereo pair, and therefore a high disparity confidence at those points.

After the ASM fitting, we obtain the 2D coordinates of n face feature points in right $R = \{(x_i, y_i), i \in [1, n]\}$ and left $L = \{(x_i^{'}, y_i^{'}), i \in [1, n]\}$ images, which are then used to obtain the final set of 3 coordinates: $P = \{p_i(x, y, d), i \in [1, n]\}$, which is the 3D sparse representation of the face. Since we use a calibrated system and rectified stereo pairs, the y coordinates of each corresponding points are then normalized to their mean. The disparity d of the points is calculated using the Euclidian distance as follows:

$$d(p_i) = \sqrt{(x_i - x_i^{'})^2 - (y_i - y_i^{'})^2}. \tag{2}$$

3.2. Dense disparity calculation

In this step, we calculate disparity of non-characteristic points of the face in a two-step process.

In the first step, we project the obtained $3_coordinates$ set P in the 3D space : $(o, \vec{x}, \vec{y}, \vec{d})$ to obtain a sparse 3D representation of the face. Then a process consisting of decomposing the face disparity into different ranges is performed. Using the assumption of the depth face symmetry and smoothness, we decompose the 3D surface with a set of level planes with associated disparity values as in Figure 1.

The decomposition step guaranties the smoothness of the final estimated disparity map and also limits the area search from the entire epipolar line to just a small segment. It also reduces the number of spikes since it limits the disparity range of each face part into the minimum and the maximum of the neighbor ranges.

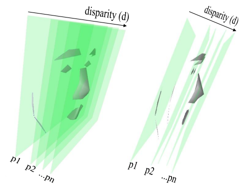

Fig. 1. The face surface decomposition.

In order to obtain the disparity of a given point p belonging to a given plane P_n, we define the disparity interval as $[DispMin_p, DispMax_p]$ where $DispMin_p$ is the disparity range associated to the plane P_n and $DispMax_p$ is that of P_{n+1}.

In the second step, we calculate the disparity of all non-characteristic points, using their disparity ranges to initialize the algorithm of the correlation. Given a face point p, with both the right projection p_r and the left projection p_l , a correlation window w and a disparity interval $[DispMin_p, DispMax_p]$, we aim at obtaining the disparity $d \in [DispMin_p, DispMax_p]$, which maximizes the correlation equation $E(d)$:

$$E(d) = Similarity(p_l(x, y), p_r(x + d, y)) \tag{3}$$

For the similarity function, we have used the SAD measure that is calculated by subtracting pixel grey level values within an $n * m$ rectangular neighborhood window w, between the reference image I_l and the target image I_r, followed by the aggregation of absolute differences within the square window.

$$\begin{aligned} SAD_{I_l(x,y),I_r(x',y'))} = \sum_{u=0}^{m}\sum_{v=0}^{n} |I_l(x+u, y+v) \\ - I_r(x^{'} + u + d, y^{'} + v)|. \end{aligned} \tag{4}$$

4. EXPERIMENTS, RESULTS AND DISCUSSION

In this section, we describe the implementation of our method, experiments and we evaluate our results quantitatively and a qualitatively.

In Figure 2, we compare the disparity map estimated by the SAD method to our method that includes the prior knowledge by applying ASM. The stereo pair is captured using a Bumblebee stereoscopic system composed of two CDD pre-calibrated cameras mounted on a horizontal support.

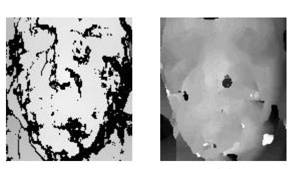

Fig. 2. Disparity map (11*11) window :(a) SAD correlation-based method, (b)Our method.

Results show that considering the face shape and its properties can enhance the disparity map in terms of smoothness and also in terms of reducing the noise (holes and spikes) occurring due to insufficient texture in homogenous face areas.

In order to compare the 3D face models generated by our method to actual 3D models, we built a binocular stereo database of 110 faces from the Texas 3D Face Database [15]. The depth map is generated using Equation 1. They are of size $325 * 488$ pixels with a resolution of 0.32 mm along the x, y, and z dimensions. A preprocessing step consisting of a median filtering is applied in order to fill holes and delete spikes. Figure 3 shows the reconstructed and the real models of some faces.

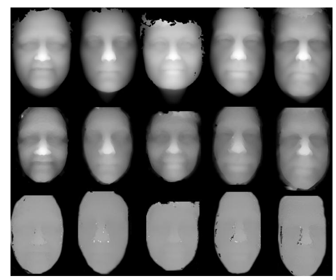

Fig. 3. Top row : original 3D model, Bottom row : depth map generated with our method.

We can observe that result of correlation based method are very poor in term of depth information and the depth rang estimated is very small comparing to the results of our method which are very similar to the original 3D model. Noises in our results were very small so they are all omitted by the median filter. However, result from correlation based method were very noisy and they still even after preprocessing.

In order to better evaluate the results we have applied the ICP (Iterative Closest Point) algorithm between original 3D models and those generated with our method is performed. Figure 4 shows the similarity matrix of this experiment and the mean of each diagonal.

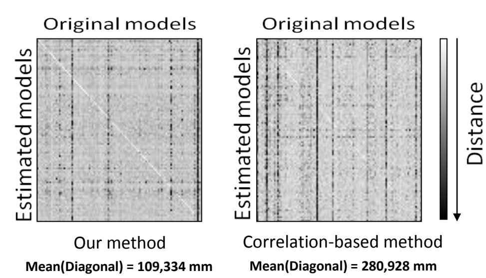

Fig. 4. Similarity matrix of the original and reconstructed 3D faces of the database.

We can see in Fig 4 that the matrix diagonal of our method is more clair then the correlation-based method were the diagonal is in clair just for some models, which demonstrates the accuracy of our result. An other important point is that the difference between a person model generated by our method and the others person models is large. However, using the correlation-based methods even if it gives a small distance between models of the same person, it gives the same with different person, which shows the specificity of the reconstructed model of our method which is garanteed by the using of the ASM of each person when calculating the depth map. The results show that the proposed strategy is robust and accurate.

5. CONCLUSIONS

This paper presents an original attempt of face depth estimation in a passive stereoscopic system. Unlike other general methods used for disparity calculation for any object, we introduced a dedicated method for face depth estimation that uses the shape characteristics of the human face, obtained by adjusting an ASM, to improve result of the correlation-based method. Our method enhanced the classical correlation-based method for disparity calculation, in terms of depth estimation efficiency, while satisfying the real-time constraint. The experimental results show that the proposed algorithm produces a smooth and dense 3D point cloud model of human face, applicable to a wide real time range of 3D face reconstruction. Our approach also opens up many perspectives for improvement and expansion. The estimation of the sparse disparity can be improved by using Active Appearance Models [16] instead of ASM, which would give more successful adjustment, because they use the texture information. The 3D Active Appearance Models [17] could also enhance the result since is robust to pose variation.

6. REFERENCES

[1] Shu-Fan Wang and Shang-Hong Lai, "Reconstructing 3d face model with associated expression deformation from a single face image via constructing a low-dimensional expression deformation manifold," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 10, pp. 2115 –2121, oct. 2011.

[2] Luuk Spreeuwers, "Fast and accurate 3d face recognition," *International Journal of Computer Vision*, vol. 93, pp. 389–414, 2011, 10.1007/s11263-011-0426-2.

[3] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust Multiview Stereopsis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 8, pp. 1362–1376, Aug. 2010.

[4] Jongmoo Choi, G. Medioni, Yuping Lin, L. Silva, O. Regina, M. Pamplona, and T.C. Faltemier, "3d face reconstruction using a single or multiple views," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, aug. 2010, pp. 3959 –3962.

[5] Amit K. Roy Chowdhury and Rama Chellappa, "Face reconstruction from monocular video using uncertainty analysis and a generic model," *Computer Vision and Image Understanding*, vol. 91, pp. 188–213, 2003.

[6] C.K. Chow and S.Y. Yuen, "Recovering shape by shading and stereo under lambertian shading model," *International journal of computer vision*, vol. 85, no. 1, pp. 58–100, 2009.

[7] Daniel Scharstein and Richard Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision*, vol. 47, pp. 7–42, April 2002.

[8] Vuong Le, Hao Tang, Liangliang Cao, and T.S. Huang, "Accurate and efficient reconstruction of 3d faces from stereo images," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pp. 4265 –4268.

[9] Unsang Park and Anil K. Jain, "3d face reconstruction from stereo video," in *Computer and Robot Vision, 2006. The 3rd Canadian Conference on*, p. 41.

[10] Ying Zheng, Jianglong Chang, Zhigang Zheng, and Zengfu Wang, "3d face reconstruction from stereo: A model based approach," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, 16 2007-oct. 19 2007, vol. 3, pp. III –65 –III –68.

[11] R. Lengagne, P. Fua, and O. Monga, "3d stereo reconstruction of human faces driven by differential constraints," *Image and Vision Computing*, vol. 18, no. 4, pp. 337–343, 2000.

[12] J.E. Cryer, P.S. Tsai, and M. Shah, "Integration of shape from shading and stereo* 1," *Pattern recognition*, vol. 28, no. 7, pp. 1033–1043, 1995.

[13] Yong Seok Heo, Kyoung Mu Lee, and Sang Uk Lee, "Robust stereo matching using adaptive normalized cross-correlation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 807–822, 2011.

[14] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," *Computer Vision–ECCV 2008*, pp. 504–513, 2008.

[15] S. Gupta, K.R. Castleman, M.K. Markey, and A.C. Bovik, "Texas 3d face recognition database," in *Image Analysis & Interpretation (SSIAI), 2010 IEEE Southwest Symposium on*. IEEE, 2010, pp. 97–100.

[16] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 6, pp. 681–685, 2001.

[17] Jing Xiao, Simon Baker, Iain Matthews, and Takeo Kanade, "Real-time combined 2d+3d active appearance models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2004, vol. 2, pp. 535 – 542.